

METHOD, SYSTEM AND COMPUTER PROGRAM PRODUCT
FOR KEEPING FILES CURRENT

BACKGROUND OF THE INVENTION

5 1. Field of the invention:

10 The present invention relates generally to data processing and in particular to file management. Still more particularly, the present invention relates to a method, system and computer program product for keeping files current.

2. Description of related art:

15 "Network computing" in a literal sense means an environment wherein a number of computers and/or peripheral devices are connected together by a communication medium (whether it be a wired or wireless medium). Additionally, the term "network" also means a communication network for transmitting data between devices that are connected to the network, such as computers, printers, storage devices and the like. There
20 are diverse forms of networks that range from a local area type, such as a local area network (LAN), to a wide area type such as a public switched telephone network (PSTN) and further to the "Internet" that has grown to a large collection of global networks as a result of
25 interconnecting respective servers.

30 A LAN is a smallest unit of a network, which is autonomously operated/managed by an independent organization, such as a college or research institution to cover a relatively narrow area, e.g., a single campus or the like. Supported with the price reduction of communication equipment reflecting the evolution of

semiconductor technologies and the enhanced functions of communication software, LANs have been primarily used in areas, such as in the research/development arenas, for the purpose of sharing computer resources, sharing/distribution of information and the like.

Wide area networks (WANs), on the other hand, are, in a simplistic sense, a larger collection of LANs wherein the servers that service each individual LAN are interconnected to create a larger network environment. Thus the services, e.g., sharing/distribution of information, are made available on a much larger global arena.

The emergence of wide area network systems, such as the Internet, has increased the accessibility of information. Connected users within these network systems have access to useful information that is made publicly available from locations, or sites, such as research facilities and libraries. These publicly available information are typically downloaded by a user, e.g., in the form of ZIP and PDF files, that are then saved on the user's memory storage devices, e.g., hard disk drive and writeable CDROM.

The files that have been downloaded may typically reside in the user's memory devices for extended periods of time prior to the information contained in those files being accessed by the user. During this extended period of time, which may be months or years, the information may become outdated or updates may exist that correct errors that have been identified in the version of the file that was downloaded. Furthermore, with the passage of time, the user may not remember the location from where the file originated and determining that location may be a difficult, if not impossible, task if the user

decides to check for a updated or newer version.

09406435.092799

SUMMARY OF THE INVENTION

5 It is therefore an object of the present invention to provide a method, system and computer program product for keeping files within a data processing system current.

10 To achieve the foregoing object, and in accordance with the invention as embodied and broadly described herein, a method, system and computer program product are disclosed for keeping files current for use in a computer system coupled to a network. The method includes: (1) evaluating a downloaded file from a source within the network to determine if a source identifier is present in the downloaded file, (2) checking the source periodically utilizing the source identifier to determine if a newer version of the downloaded file exists and (3) replacing, in response to the presence of a newer version of the downloaded file, the downloaded file with the newer version. The method further includes attaching, in response to the source identifier not being present, a source descriptor to the downloaded file.

15 20 In one embodiment of the present invention, the step of replacing the downloaded file includes the steps of (1) providing an indication to a user that the newer version of the file exists, (2) prompting the user to replace the downloaded file with the newer version and (3) replacing, in response to the user requesting the newer version, the downloaded file with the newer version.

25 30 In another embodiment of the present invention, the source identifier is located in the extended attribute of the downloaded file. It should be noted, however, that the location of the source identifier may vary depending

on the type of file format or operating system employed.

5 In yet another embodiment of the present invention, the downloaded file is a PDF file. Alternatively, in another advantageous embodiment, the downloaded file is a ZIP file. It should be readily apparent to those skilled in the art that the present invention may be advantageously practiced with other file format methodologies.

10 In another embodiment of the present invention, the step of checking the source periodically includes defining a time interval. In one advantageous embodiment, the time interval is user defined. Alternatively, the step of checking the source may be accomplished whenever the downloaded file is opened or "on-demand" by a user.

15 In one embodiment of the present invention, the network is a packet network. Of course, the present invention may also be advantageously practiced in other network environments such as local area networks (LANs) and wide area networks (WANs). The present invention does not contemplate limiting its use to any one particular network environment.

20 The foregoing description has outlined, rather broadly, preferred and alternative features of the present invention so that those skilled in the art may better understand the detailed description of the invention that follows. Additional features of the invention will be described hereinafter that form the subject matter of the claims of the invention. Those skilled in the art should appreciate that they can readily use the disclosed conception and specific embodiment as a basis for designing or modifying other

25

30

structures for carrying out the same purposes of the present invention. Those skilled in the art should also realize that such equivalent constructions do not depart from the spirit and scope of the invention in its broadest form. \

09406435.092799

BRIEF DESCRIPTION OF THE DRAWINGS

For a more complete understanding of the present invention, reference is now made to the following descriptions taken in conjunction with the accompanying drawings, in which:

FIGURE 1 illustrates an exemplary network system that provides a suitable environment for the practice of the present invention;

FIGURE 2 illustrates an embodiment of a controller employing a file updating system constructed utilizing the principles disclosed by the present invention; and

FIGURE 3 illustrates a high level logic flow diagram of an embodiment of a file updating process utilizing the principles disclosed by the present invention.

09406435-092799

DESCRIPTION OF THE ILLUSTRATIVE EMBODIMENTS

With reference now to the figures, and in particular, with reference to **FIGURE 1**, there is depicted an exemplary network system **100** that provides a suitable environment for the practice of the present invention.

Network system **100** includes a computer system **110**, such as a personal computer (PC), that is coupled to first and second sites **120**, **130**, respectively, via a packet network **140**, e.g. the Internet. It should be noted that the present invention may also be advantageously practiced in other network environments, such as a local area network (LAN). First and second sites **120**, **130** are generally sites that provide information to users, such as libraries and research facilities, that are connected to network system **100**. First and second sites **120**, **130** typically provide services, which may be free, i.e., no monetary charges are required to access the site services, that include application programs, such as Acrobat reader from Adobe. These "free" programs are generally available in a file that a user, such as computer system **110**, would download through packet network **140** to a memory device (not shown), such as a hard disk or a writeable CDROM, coupled to computer **110**.

As discussed previously, the files that have been downloaded may typically reside in the user's memory devices for extended periods of time prior to the information contained in those files being accessed by the user. During this extended period of time, which may be months or years, the information may become out dated or updates may exist that correct errors that have been identified in the version of the file that was downloaded. Furthermore, with the passage of time, the user may not remember the location from where the file originated and determining that location may be a

difficult, if not impossible, task if the user decides to check for a updated or newer version.

Referring now to **FIGURE 2**, there is illustrated an embodiment of a controller 200 employing a file updating system constructed utilizing the principles disclosed by the present invention. Controller 200 (analogous to computer system 100 illustrated in **FIGURE 1**), in an advantageous embodiment, is a personal computer manufactured by IBM Corporation of Armonk, N.Y. It should also be readily apparent to those skilled in the art, however, that alternative computer system architectures may be employed. Generally, controller 210, embodied in a PC computer, comprises a bus 215 for communicating information, a processor 220 coupled to bus 215 for processing information, a random access memory (not shown) coupled to bus 215 for storing information and instructions for processor 220, a read-only memory (not shown) coupled to bus 215 for storing static information and instructions for processor 220, a display device 250 coupled to bus 215 for displaying information for a computer user, an input device (not shown) coupled to bus 215 for communicating information and command selections to processor 220 and a data storage device (not shown), such as a magnetic disk and associated disk drive, coupled to bus 215 for storing information and instructions.

Processor 220 may be any of a wide variety of general purpose processors or microprocessors, such as the i486™ or Pentium™ brand microprocessor manufactured by Intel Corporation of Santa Clara, California. However, it should be apparent to those skilled in the art that other varieties of processors may be utilized in a computer system. Display device 250 may be a liquid crystal device, cathode ray tube (CRT), or other suitable

display device. The data storage device may be a conventional hard disk drive, floppy disk drive, or other magnetic or optical data storage device for reading and writing information stored on a hard disk drive, floppy disk drive, or other magnetic or optical data storage medium.

In general, processor 220 retrieves processing instructions and data from a data storage medium using the data storage device and downloads this information into random access memory for execution. Thereafter, processor 220 then executes an instruction stream from random access memory or read only memory. Command selections and information input at the input device are used to direct the flow of instructions executed by processor 220. The results of this processing execution are then displayed on display device 250.

Controller 210 further includes an update manager 230 that is coupled to processor 220. Update manager 230, in an advantageous embodiment, is embodied as a set of computer executable instructions stored on a computer readable medium, such as the hard disk. It should be readily apparent, however, to those skilled in the art that update manager 230 may also be implemented in hardware, firmware, software and any combination thereof. The present invention does not contemplate limiting its practice to any particular form of implementation.

Referring now to **FIGURE 3**, with continuing reference to **FIGURES 1 and 2**, depicted is a high level logic flow diagram of an embodiment of a file updating process 300 utilizing the principles disclosed by the present invention. Process 300 begins, as depicted in step 310, when the process is queued for execution. Next, as illustrated in step 320, controller 210 selects and

downloads a file from a source site, e.g., first or second sites 120, 130. In an advantageous embodiment, the downloaded file may be a PDF file. Alternatively, in another advantageous embodiment, the downloaded file is a ZIP file. It should be noted, however, that the practice of the present invention is not limited to any particular type of file format methodology.

Following the downloading of the selected file, update manager 230 evaluates the downloaded file to determine if the file has a source identifier associated with the site from which it was obtained, as depicted in decisional step 330. The evaluation is accomplished, e.g., by looking at the extended attributes or directory in the downloaded file, to see if an identifier, such as an uniform resource locator (URL), associated with the site is present. If it is determined that there is no source identifier associated with the downloaded file, update manager attaches a source descriptor to the downloaded file, as illustrated in step 340. The attachment or "tagging" the source descriptor may be accomplished, e.g., by adding a new comment entry in a ZIP file. Generally, most relevant file formats have room for additional comment text or other attribute string. This attribute string, i.e., "source identifier," is added to the file to identify the source location of the file to which it is attached or in which it appears. It should be noted that certain operating systems, such as OS/2, support extended attributes that are associated with a file. Therefore, the source identifier may be stored as an extended attribute and does not need to be inserted inside the file. The source identifier, in an advantageous embodiment, may contain the following:

- (1) A signature string that is unlikely to appear in

any other portion of the file. This signature string is used to find the source identifier within the file.

5 (2) A URL or other locator string that identifies the location from which the file (its newest version) can be retrieved.

(3) A date/time and version number corresponding to the file.

10 (4) A checksum string covering the prior entries to make it less likely that random data content would be mistaken for a signature string.

15 When the source identifier is located within the file, it should be also located as far towards the end of the file as possible, so that the last signature string in the file is the one that is a part of the source identifier. In the event that an uncompressed archive file, such as a ZIP file, contains other ZIP files with their own source identifiers, locating the source identifier at the end of the file would prevent the present invention from incorrectly using an earlier embedded source identifier in the file. It should be
20 readily apparent to those skilled in the art that the preferred location of the source identifier is different for different file types and depends on the methodology employed by the file to contain comment strings.

25 *Sub 217* Alternatively, in another advantageous embodiment, an entry may be entered in a specially coded file registry associated within controller 210 that records, at a minimum, the name of the downloaded files and a source descriptor identifying the originating source,
30 such as an URL, from where the downloaded file was obtained. Other entries in this file registry may

5 include a time stamp of when the downloaded file was
retrieved from the source site. In the case where
controller 210 is running an OS/2 operating system, the
source descriptor information may be stored in the
extended attributes of the downloaded file. It should
10 readily apparent to those skilled in the art that the
location where the source descriptor is stored is
dependent on the file format methodology employed by
controller 210, e.g., for a PDF file, the source
descriptor information may not be added to the file in a
manner that may disable older versions of Acrobat
viewers. In the case of a PDF file or with the OS/2
operating system discussed above, in an advantageous
embodiment, the source descriptor could be stored within
15 the file by replacing a text of a comment or any other
embedded string in the document file with a specially
coded string that has an unique digital signature.

20 *sub 300* Following the attachment of the source descriptor,
process 300 enters a "dormant" or waiting period, as
depicted in step 345, until such time as when the
downloaded file is opened by a user or, in another
advantageous embodiment, at a predetermined time
interval. The time interval is typically set by the user
and may be programmed to be as short as daily or longer
as every six months. Alternatively, in another
embodiment, a triggering event may be a "on-demand"
request by the user to update the file. Similarly, if it
is determined in step 330 that the downloaded file has a
source identifier included in it, process 300 proceeds to
30 wait until the downloaded file is opened by the user or
at the predetermined time interval.

In the event that the downloaded file is opened by
the user, or in another alternative embodiment, at the
expiration of the predetermined time interval, update

manager 230 retrieves the source identifier from the downloaded file and proceeds to check the file's source site to determine if a newer version of the file is present, as depicted in decisional step 350.

5 If there is a newer version of the file present, the update manager 210 proceeds to provide an indication to the user that a newer version of the file is available. The user may be prompted with a message, such as "Updated version of file available, would you like to replace existing file Y/N" displayed on display 250. If the user
10 responses with an affirmative Y, update manager 230 replaces the "older" file with its newer or updated version as illustrated in step 260. Alternatively, in another embodiment, replacing the older file involves
15 renaming the older file. With this approach, the older version of the file is still available along with the newer, i.e., most current, file version. After update manager has replace the file with its newer version or if
20 the user had decided that the newer version is not desired, process 300 does nothing and returns to its dormant state, i.e., step 345, to wait for the next triggering event, e.g., when the file is opened again, on-demand by the user or at the end of the next time interval.

25 The present invention provides for the automatic updating of files that have been downloaded from a source site that is coupled to a user's system without the user having to remember where the file was obtained from. Consequently, with an attached source identifier, any
30 downloaded file can be traced to its originating source site. Furthermore, mirror sites can perform automatic file updates based on the file content rather than using a separate directory of file locations.

It should be noted that although the present invention has been described in the context of a computer system, those skilled in the art will readily appreciate that the present invention is also capable of being distributed as a computer program product in a variety of forms; the present invention does not contemplate limiting its practice to any particular type of signal-bearing media, i.e., computer readable medium, utilized to actually carry out the distribution. Examples of signal-bearing media includes recordable type media, such as floppy disks and hard disk drives, and transmission type media such as digital and analog communication links.

In a preferred embodiment, the present invention is implemented in a computer system programmed to execute the method described herein. Accordingly, in an advantageous embodiment, sets of instructions for executing the method disclosed herein are resident in RAM of one or more of computer systems configured generally as described hereinabove. Until required by the computer system, the set of instructions may be stored as computer program product in another computer memory, e.g., a disk drive. In another advantageous embodiment, the computer program product may also be stored at another computer and transmitted to a user's computer system by an internal or external communication network, e.g., LAN or WAN, respectively.

The present invention may be embodied in other specific forms without departing from its spirit or essential characteristics. The described embodiments are to be considered in all respects as illustrative and not restrictive. The scope of the invention is, therefore, indicated by the appended claims rather than by the foregoing description. All changes which come within the

meaning and range of equivalency of the claims are to be embraced within their scope.

09406435, 092799